Microdata and TableBuilder: Qualifications and work

Contains TableBuilder and DataLab products

On this page

[Accessing the data](#)
[Data and file structure](#)
[Using TableBuilder](#)
[Using DataLab](#)
[Data downloads](#)
[Methodology](#)
[Post release changes](#)

Accessing the data

The Qualifications and work (Q&W) publication presents detailed information about the educational history of people and the relevance of each qualification to their working lives.

Data collected for up to five qualifications includes the level and field, year of completion and whether the qualification was attained in Australia. Further information was collected on incomplete qualifications, main language spoken at home, and other demographics. See Qualifications and work (https://www.abs.gov.au/statistics/people/education/qualifications-and-work/latest-release) for summary results, methodology and other information.

The following microdata products are available from this survey:

- TableBuilder - produce your own tables and graphs. TableBuilder is available for the following survey years: 2010/11*, 2015, 2018/19, and 2022/23.
- DataLab - expanded microdata is available in DataLab for the following survey years: 2018/19.

Compare data services (https://www.abs.gov.au/about/data-services/compare-data-services) to see what's right for you or Frequently asked questions (https://www.abs.gov.au/statistics/microdata-tablebuilder/frequently-asked-questions) .

*Survey was called 'Learning & Work' (L&W) in 2010/11

Data and file structure

Data available by level

The Qualifications and work microdata is available across two levels.

1. Person Level
2. Qualification Level

Feedback

The two levels are hierarchical. The first level relates to people and the second level relates to the qualifications of those people. Visually, the person level file has one row of data for each respondent, containing all the person level data items. The qualification level contains up to five rows for each respondent, one for each of their qualifications.

The person level contains characteristic data such as age, sex, marital status, employment, and educational status including the number of non-school qualifications completed and the level and field of highest educational attainment. This level also has information about the person's household, for example, the number of persons who are usual residents in the household and the overall household income. In addition, the person level includes geographic data items such as state or territory of usual residence.

The qualification level contains details about each completed qualification that a respondent has reported. The type of detail available about each qualification include the level and field of the qualification, the year each qualification was completed, whether the qualification was completed in Australia and if employed, the relevance of each qualification to the person's current job.

A complete data item list can be accessed from the Data downloads section.

Unit identifiers

Every record on each level of the detailed microdata files are uniquely identified. For more information about the identifiers, see the Weights and/or Identifiers section of the relevant data item list.

Using TableBuilder

Please refer to relevant sections in the TableBuilder (https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder) main page for information about how to create basic tables, custom groups, graphs and large tables. Information specific to the Q&W TableBuilder, which enables users to understand, interpret and tabulate the data, outlined below.

Weights

Weighting is the process of adjusting results from a sample survey to infer results for the total population. To do this, a 'weight' is allocated to each record. The weight is the value that indicates how many population units are represented by each sample unit.

To produce estimates for the in-scope population you must use weight fields in your tables. In TableBuilder they can be found under the 'Summation Options' category in the left hand pane. If you do not select a weight field, TableBuilder will apply the person weight by default. This will give you estimates of the number of persons. To produce estimates of the number of qualifications you would have to add the qualification weights to your table.

Note that in the 2018/19 TableBuilder, the person level weights are labelled as 'Persons' and are located under 'Person Level Weights'. In the 2022/23 TableBuilder, the person level weights are labelled as 'Persons (3)' where the 3 represents Person counts. Similarly, the qualification weights in the 2018/19 TableBuilder are labelled as 'Qualifications' and are located under 'Qualification Level Weights'. In the 2022/23 TableBuilder, the qualification level weights are labelled as 'Persons (6)' where 6 represents Qualification counts. This is purely a labelling difference only and is also described in the Q&W 22/23 TableBuilder data item list (under 'Weights' tab).

If you are estimating the number of persons with certain characteristics (e.g. 'Number of non-school qualifications completed') the person level weight must be used.

To estimate the number of qualifications (e.g. 'Number of non-school qualifications completed') the qualification level weight must be used.

Qualification level data items are weighted according to the characteristics of the person who undertook the

qualification, and therefore the weights for each qualification are the same as the weight for the person. For example, if a person in the sample has a weight of 600 and that person has completed three non-school qualifications then the person represents 600 people in the total population and 1,800 qualifications.

Cross-tabulating data items on the same level

Cross-tabulating data from the person level with other data items from the same level will produce data about people. For example, cross-tabulating the geographic variable 'State or territory of usual residence' by the 'Level of most recent non-school qualification (ASCED)' produces a table showing the number of people in each region by the most recent qualification they have obtained.

Cross-tabulating data from the qualification level with other data items from the same level will produce data about qualifications when using the qualification weight. For example, cross-tabulating 'Level of non-school qualification (ASCED) - Qualification level' by 'Whether completed non-school qualification through an Australian institution - Qualification level' produces a table showing the number of qualifications completed through an Australian institution. If a respondent has several qualifications, each of those qualifications is included in the table. If the same cross-tabulation is generated but using the person weight instead of the qualification weight, it produces a table showing the number of people who completed a non-school qualification through an Australian institution.

Using qualification ordering data items and flags

To assist with analysis, several variables have been created to help isolate and order qualifications. These are located under 'Five Highest Non-School Qualifications' and 'Qualification Details and Relevance to Current Job' in the left hand pane.

By using a qualification level ordering data item (e.g. 'Highest to fifth highest non-school qualification' and selecting the 'Highest qualification' category) only one qualification for each respondent is included in a table. Selecting either the person weight or the qualification weight when using a qualification ordering data item or flag will produce essentially the same result, any difference being the result of perturbation acting slightly differently when using the different weights.

Cross-tabulating person level by qualification level data items

Cross-tabulating data items from the person level with data items from the qualification level can produce data about people or qualifications depending on the weight being used. Caution should be used as a person with multiple qualifications may have the same qualifications counted only once in a table.

Using a qualification ordering data item or flag (as described in 'Using Qualification ordering data items and flags' above) may be worthwhile when cross-tabulating person level with qualification level data items as only one selected qualification will be included in the tabulation.

Cross-tabulating qualification level data items by person level data items using the person weight - When using a qualification ordering data item or flag

When using a qualification ordering data item (e.g. 'Highest to fifth highest non-school qualification') and selecting the 'Highest qualification' category' in a table that cross-tabulates a qualification level data item by a person level data item, either the person or the qualification weight can be used and the same output will be generated, with any difference being due to perturbation. Restricting the table to a single qualification (e.g. highest non-school qualification) for each person in effect turns this into a person level data item, as TableBuilder only needs to read one row of data from the qualification level for each person.

Cross-tabulating qualification level data items by person level data items using the person weight - When NOT using a qualification ordering data item or flag

When a qualification ordering data item is not used, TableBuilder will read each row of data from the qualification level for each person. In this case, TableBuilder effectively calculates the tabulation as a 'multi-response' table (i.e. the same person can be counted more than once), but it counts the same categories of information about different qualifications only once. It treats them as 'one or more occurrences' of that category. For example, if a respondent completed three qualifications, and the respondent is currently working in the field of one of these qualifications but not the other two, then the person would be counted in each column of the data item 'Currently working in job in the same field as main field of non-school study - Qualification level'.

Therefore, in these particular types of tabulations, components of the table will not add to the total number of persons (as persons can be counted more than once), but the total will be the correct count of persons as TableBuilder calculates the total in such a way that each person is only counted once. An example table below (without a qualification ordering data item) for 'Currently working in job in the same field as main field of non-school study - Qualification level' shows results for all qualifications for a person, so they can appear in both columns:

Persons ⓘ		Persons #		
Age of person (10 year age groups) ⓘ Σ C		age 15-64		
Currently working in the same field as main field of non-school qualification - Qual level ⓘ Σ C		Currently working in the same field as main field of study of qualification	Not currently working in the same field as main field of study of qualification	Total
<< Labour force status ⓘ Σ C		Employed	Employed	Employed
Level of non-school qualification (ASCED) - Qualification level ⓘ Σ C	⇕	⇕	⇕
Postgraduate Degree		753.1	274.1	995.2
Graduate Diploma and Graduate Certificate		524.5	190.5	683.9
Bachelor Degree		2,507.8	1,330.9	3,701.9
Advanced Diploma and Diploma		1,111.2	761.5	1,805.6
Certificate III and IV		2,139.6	1,276.7	3,21
Certificate I and II		212.2	245.3	43
Certificate n.f.d.		79.4	42.4	11
Level not determined		340.5	197.2	52
Total		6,172.5	3,647.9	8,44

▼ Annotation Descriptions
(000's)

The following table shows the same data item for the 'Highest qualification' only by using the ordering data item 'Highest to fifth highest non-school qualification'. In this case, people only appear once only in either column and consequently columns add to totals (taking into account small differences in perturbation).

Persons ⓘ		Persons #		
Age of person (10 year age groups) ⓘ Σ C		age 15-64		
Currently working in the same field as main field of non-school qualification - Qual level ⓘ Σ C		Currently working in the same field as main field of study of qualification	Not currently working in the same field as main field of study of qualification	Total
<< Labour force status ⓘ Σ C		Employed	Employed	Employed
Highest to fifth highest non-school qualification ⓘ Σ C		Highest qualification	Highest qualification	Highest qualification
Level of non-school qualification (ASCED) - Qualification level ⓘ Σ C	⇕	⇕	⇕
Postgraduate Degree		740.3	255.7	995.2
Graduate Diploma and Graduate Certificate		418.0	127.8	545.0
Bachelor Degree		1,716.9	840.9	2,554.6
Advanced Diploma and Diploma		794.2	555.3	1,345.7
Certificate III and IV		1,643.2	832.5	2,473.7
Certificate I and II		112.3	108.4	223.7
Certificate n.f.d.		36.7	* 13.9	52.2
Level not determined		165.1	89.0	255.0
Total		5,619.0	2,823.5	8,443.2

▼ Annotation Descriptions
(000's)

In summary, qualification level data items can be cross-tabulated with person level data items with or without qualification ordering data items or flags. Qualification ordering data items and their relevant category should be included in tables when a user wants information only about one particular qualification (e.g. the highest

qualification or the most recent qualification), but should not be used in tables looking at all qualifications.

Multi-response data items

A number of data items allow respondents to report more than one response. These are referred to as 'multi-response data items'. An example of a multi-response item in the Q&W TableBuilder is 'All sources of personal income (multiple response)' located under 'Income' in the left hand pane.

When a multi-response data item is tabulated, a person is counted against each response they have provided (e.g. a person who responds 'employee income' and 'unincorporated income' and 'government pensions and allowances' will be counted once in each of these three categories).

As a result, each person in the appropriate population is counted at least once, and some persons are counted multiple times. Therefore, the total for a multi-response data item will be less than or equal to the sum of its components.

Not applicable categories

Most data items included in the TableBuilder file include a 'Not applicable' category. The classification values of these 'Not applicable' categories, where relevant, are shown in the TableBuilder data item list. The 'Not applicable' category generally represents the number of people who were not asked a particular question or the number of people excluded from the population for a data item when that data was derived (e.g. Number of non-school qualifications completed is not applicable for people who have not completed a non-school qualification).

Table populations

The population relevant to each data item is identified in the data item list and should be kept in mind when extracting and analysing data. The actual population estimate for each data item is equal to the total cumulative frequency minus the 'Not applicable' category.

The purpose of the population data item 'Non-Indigenous flag' is to assist users in producing non-indigenous data only. It should not be used to estimate for the Indigenous population through differencing, as the scope of Q&W excludes persons living in Indigenous communities.

Continuous data items

The Q&W TableBuilder includes several continuous variables:

- They can have a response value at any point along a continuum.
- Some continuous data items are allocated special codes for certain responses (e.g. 000 = 'Not applicable').
- When creating ranges in TableBuilder for such continuous items, special codes will automatically be excluded. Therefore the total will show only 'valid responses' rather than all responses (including special codes). These codes are shown in the data item list.
- Continuous items with special codes have a corresponding categorical item on the Person or Qualification level that provides the ability to display data for the special code. Refer to the data item list.

Confidentiality

A confidentiality process called perturbation is applied to the data in TableBuilder to avoid releasing information that may lead to the identification of individuals, families, households, dwellings or businesses. See Confidentiality and relative standard error (https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder/confidentiality-and-relative-standard-error) .

Using DataLab

The DataLab environment allows real time access to detailed microdata from the Q&W survey. The DataLab is an interactive data analysis solution available for users to run advanced statistical analyses, for example, multiple regressions and structural equation modelling. The DataLab environment contains recent versions of analytical software, including R, SAS, Stata and Python. Controls in the DataLab have been put in place to protect the identification of individuals and organisations. All output from DataLab sessions is cleared by an ABS officer before it is released.

For information about all of the data items available in the DataLab please see the Microdata data item list in the Data downloads section.

For more information, including prerequisites for DataLab access, please see the [DataLab (https://www.abs.gov.au/statistics/microdata-tablebuilder/datalab)](https://www.abs.gov.au/statistics/microdata-tablebuilder/datalab) page.

DataLab files, data items and Identifiers

There are two Q&W Datalab files; a Person level file and a Qualification level file.

The Person level contains a range of data items relating to the respondent, including; demographic, employment, income, education, study and geographical details.

The Qualification level contains specific details on each of the five highest qualifications that the respondent has obtained. The file is hierarchical with a unique record (row) for each of their qualifications.

Every record on each file is uniquely identified. These Identifiers can be used to copy information from one level of the file to another. For example, demographic detail from the person level, such as age and sex, can be added to the qualification level (in a one to many merge) using the identifiers. For information about all of the data items available and identifiers please see the Microdata data item list in the Data downloads section.

DataLab test file

A test file is available in the Data downloads section for researchers to become familiar with the data structure and prepare code/programs before applying for or beginning a DataLab session.

Person weights, qualification weights and reliability of estimates

As the survey was conducted on a sample of households in Australia, it is important to take account of the method of sample selection when deriving estimates from the detailed microdata. This is important as a person's chance of selection in the survey varied depending on the state or territory in which the person lived. If these chances of selection are not accounted for by use of appropriate weights, the results could be biased.

Each person record has a main weight (FINWTPQ). This weight indicates how many population units are represented by the sample unit. When producing estimates of sub-populations from the detailed microdata, it is essential that they are calculated by adding the weights of persons in each category and not just by counting the sample number in each category. If each person's weight were to be ignored when analysing the data to draw inferences about the population, then no account would be taken of a person's chance of selection or of different response rates across population groups, with the result that the estimates produced could be biased. The application of weights ensures that estimates will conform to an independently estimated distribution of the population by age, by sex, etc. rather than to the distributions within the sample itself.

It is also important to calculate a measure of sampling error for each estimate. Sampling error occurs because only part of the population is surveyed to represent the whole population. Sampling error should be considered when interpreting estimates as this gives an indication of accuracy and reflects the importance that can be placed on interpretations using the estimate. Measures of sampling error include standard error (SE), relative standard error

(RSE) and margin of errors (MoE). These measures of sampling error can be estimated using the replicate weights. The replicate weight variables provided on the microdata are labelled WPQ01XX , where XX represents the number of the given replicate group. The exact number of replicates will vary depending on the survey but will generally be 30, 60 or 200 replicate groups. As an example, for survey microdata with 30 replicate groups, you will find 30 person replicate weight variables labelled WPQ0101 to WPQ0130.

Qualifications are weighted according to the characteristics of the person who undertook the qualification, and therefore the weights for each qualification are the same as the weight for the person (FINWTPQ). For example, if a person in the sample has a weight of 600 and that person has completed three non-school qualifications then the person represents 600 people in the total population and 1,800 qualifications.

Qualification ordering data items

If you are estimating the number of persons with certain characteristics the person level file and person level weight is used.

To estimate the total number of qualifications the qualification level and qualification weight is used.

To assist with analysis, several variables have been created to help isolate and order qualification level data. These are detailed on the 'Five highest qualifications' tab of the data item list.

Using a qualification ordering data item such as 'Highest to fifth highest non-school qualification' allows for one qualification to be selected from this level (such as the highest qualification). This will allow for estimates of persons to be produced, rather than an estimate of qualifications.

In summary, qualification level data items can be cross-tabulated with person level data items with or without a qualification ordering data item. A qualification ordering data item should be included in analysis when a user wan information only about one particular qualification (e.g. the highest qualification or the most recent qualification), but should not be used in tables looking at all qualifications.

Using Replicate Weights for Estimating Sampling Error

Overview of replication methods

ABS household surveys employ complex sample designs and weighting which require special methods for estimating the variance of survey statistics. Variance estimators for a simple random sample are not appropriate for this survey microdata.

A class of techniques called 'replication methods' provide a general process for estimating variance for the types of complex sample designs and weighting procedures employed in ABS household surveys. The ABS uses a method called the Group Jackknife Replication Method.

A basic idea behind the replication approach is to split the sample into G replicate groups. One replicate group is then dropped from the file and a new set of weights is produced for the remaining sample. This is repeated for all G replicate groups to provide G sets of replicate weights. For each set of replicate weights, the statistic of interest is recalculated and the variance of the full sample statistic is estimated using the variability among the replicate statistics. The statistics calculated from these replicates are called replicate estimates. Replicate weights provided on the microdata file enable variance of survey statistics, such as means and medians, to be calculated relatively simply (Further technical explanation can be found in [Section 4 of Research Paper: Weighting and Standard Error Estimation for ABS Household Surveys (Methodology Advisory Committee (https://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.029)](https://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.029)).

How to use replicate weights

To calculate the standard error of any statistic derived from the survey data, the method is as follows:

1. Calculate the estimate of the statistic of interest using the main weight.

2. Repeat the calculation above for each replicate weight, substituting the replicate weight for the main weight and creating G replicate estimates. In the example where there are 30 replicate weights, you will have 30 replicate estimates.

3. Use the outputs from step 1 and 2 as inputs to the formula below to calculate the estimate of the Standard Error (SE) for the statistic of interest.

$$SE(y) = \sqrt{\frac{G-1}{G} \sum_{g=1}^{G} (y_{(g)} - y)^2}$$

- G = Number of replicate groups

- g = the replicate group number

- $y_{(g)}$ = Replicate estimate for group g, i.e. the estimate of y calculated using the replicate weight for g

- y=the weighted estimate of y from the sample

From the replicate variance you can then derive the following measures of sampling error: relative standard error (RSE), or margin of error (MOE) of the estimate.

$$Relative\ Standard\ Error\ (RSE) = \frac{SE}{Estimate}$$

$$Margin\ of\ Error\ (MoE) = 1.96\ \times\ SE$$

An example of calculating the SE for an estimate of the mean

Suppose you are calculating the mean value of earnings, y, in a sample. Using the main weight produces an estimate of $500.

You have 5 sets of Group Jackknife replicate weights and using these weights (instead of the main weight) you calculate 5 replicate estimates of $510, $490, $505, $503, $498 respectively.

To calculate the standard error of the estimate you will substitute the following inputs into the first equation calculating the SE under 'How to use replicate weights':

G = 5

y=500

g=1, $y_{(g)}$= 510

g=2, $y_{(g)}$= 490

...

$$SE(y) = \sqrt{\frac{5-1}{5} \sum_{g=1}^{5} (y_{(g)} - 500)^2}$$

$$SE(y) = \sqrt{\frac{4}{5}((510-500)^2 + (490-500)^2 + (505-500)^2 + (503-500)^2 + (498-500)^2}$$

$$SE(y) = \sqrt{\frac{4}{5} \times 238}$$

$$SE(y) = 13.8$$

To calculate the RSE you divide the SE by the estimate of y ($500) and multiply by 100 to get a %:

Feedback

$$RSE(y) = \frac{13.8}{500} \times 100$$

$$RSE(y) = 2.8\%$$

To calculate the margin of error you multiply the SE by 1.96:

$$Margin\ of\ Error(y) = 13.8 \times 1.96$$

$$Margin\ of\ Error(y) = 27.05$$

Data downloads

Data cubes

⬇ **Download all (827.42 KB)**

Q&W 2018-19 Microdata data item list

⬇ **Download** XLS
[1.8 MB]

Q&W 2015 Microdata data item list

⬇ **Download** XLS
[565 KB]

L&W 2010-11 Microdata data item list

⬇ **Download** XLS
[275.5 KB]

Detailed Microdata test file

⬇ **Download** ZIP
[9.19 MB]

Q&W 2022-23 Microdata data item list

⬇ **Download** XLSX
[521.11 KB]

Feedback

Methodology

Show all

See Qualifications and Work, Methodology (/statistics/people/education/qualifications-and-work/latest-release#methodology) for information on:

- Data collection
- Processing the data
- Comparing the data
- Data release
- Glossary
- Abbreviations

Post release changes

[Show all](#)

13/06/2024 - TableBuilder for Q&W 2022-23 is now available.

Previous catalogue number

This release previously used catalogue number 4235.0.55.001.

Feedback